Intramolecular DNA triplexes: Unusual sequence requirements and influence on DNA polymerization

(H*-DNA/DNA polymerase/termination)

ANDREY DAYN, GEORGE M. SAMADASHWILY, AND SERGEI M. MIRKIN[†]

Department of Genetics, University of Illinois at Chicago, Chicago, IL 60612

ABSTRACT   Homopurine–homopyrimidine mirror re-
peats are known to form intramolecular DNA triplexes in vitro.
By probing with chemical agents specific for unusual DNA
conformations, we have now demonstrated the formation of
intramolecular triplexes consisting of G·G·C and T·A·T base
triplets by DNA sequences that are neither homopurine–
homopyrimidine nor mirror repeats. This finding significantly
enlarges the number of sequences that could form DNA tri-
plexes. The observed triplexes are stable under the conditions
that are optimal for DNA polymerases in vitro. We found that
triplex formation causes specific termination of DNA polymer-
ization in vitro. This effect is detected for different DNA
polymerases and may have implications for the regulation of
DNA replication in vivo.

Homopurine–homopyrimidine DNA sequences adopt un-
usual triple-helical DNA structures in supercoiled DNA (for
review, see refs. 1 and 2). Two different types of intramo-
lecular DNA triplexes are known. Under mildly acidic con-
ditions, H-DNA is formed by half of the purine strand and the
hairpin formed by the pyrimidine strand. It consists of
C·G·C⁺ and T·A·T base triads (3–10). In the presence of
bivalent cations, a related H*-DNA structure is formed by
half of the pyrimidine strand and a hairpin formed by the
purine strand. It is stabilized by G·G·C and A·A·T base triads
(11–14). To form triplexes consisting of these triads, homopu-
rine–homopyrimidine DNA sequences must be mirror re-
peats (H palindromes) (5). This was clearly shown for
H-DNA: even a single deviation from mirror symmetry
significantly disfavors triplex formation (15). For H*-DNA,
however, such analysis has not been carried out yet.

Beal and Dervan (16) recently detected the formation of
intermolecular triplexes where purine-rich oligonucleotides
bind to double-helical DNA antiparallel to the purine-rich
strand (i.e., in the same way as in H*-DNA). As expected,
G·G·C triads play the key role in these complexes. The
relative stabilities of other triads are T·A·T > A·A·T > C·A·T.
This shows that an intermolecular triplex can exist that
consists of G·G·C and T·A·T base triads. We assumed,
therefore, that an intramolecular triplex consisting of G·G·C
and T·A·T base triads could exist as well.

To form such a triplex, a sequence must conform to
unusual requirements (Fig. 1). One can see that the guanine
and thymine residues in the half of the purine-rich strand that
turns back to make a triplex form Hoogsteen hydrogen bonds
with guanine and adenine residues, respectively, in the other
half of the same strand (Fig. 1). Therefore, guanine residues
should be arranged in a mirror repeated way (as in H- or
H*-DNA), whereas thymine residues concentrated in one
half of the purine strand should be reflected by adenine
residues in the other half. As a result, such sequences would

be neither homopurine–homopyrimidine, because they con-
tain thymine residues in otherwise purine-rich strands, nor
mirror repeated, because adenine and thymine bases are
positioned in an inverted repeat way (Fig. 1). We have cloned
such sequences and analyzed their structure by using chem-
icals specific toward non-B-DNA conformations: diethyl
pyrocarbonate (DEPC), OsO₄, and chloroacetaldehyde
(CAA). We found that they indeed form H*-DNA when in the
supercoiled state and in the presence of bivalent cations.

Recently it was found that DNA polymerases in vitro
terminate in the middle of single-stranded homopurine or
homopyrimidine sequences cloned in M13 phage (17, 18). It
was suggested that this is due to triplex formation in the
course of DNA polymerization, so that the newly synthesized
DNA chain is involved in triplex formation. The correspond-
ing situation in double-stranded DNA where preexisting
DNA triplexes could provide a stronger termination signal
has never been studied. Therefore, we studied the activity of
DNA polymerases in vitro on double-helical DNA templates
containing our triplex-forming DNA sequences. We found
prominent elongation stops at locations that correlate pre-
cisely with the triplex borders as defined by chemical anal-
ysis. Thus, we suggest that DNA polymerases are unable to
effectively read through triple helices preformed in a template
DNA.

MATERIALS AND METHODS

Plasmid Construction. Oligonucleotides corresponding to
sequences I and II (Fig. 1) were synthesized and cloned in the
pUC19 polylinker between the BamHI and EcoRI sites. To
obtain single-stranded DNA, sequence II was recloned in two
orientations in the phagemid pBluescript SK(−) (Stratagene).
As a result, either the purine-rich or the pyrimidine-rich
strand could be rescued in a single-stranded state after the
isolation of phage DNA. Single-stranded DNA was isolated
as described in ref. 19.

Chemical Modification of DNA. Supercoiled DNAs were
modified in 25 mM sodium cacodylate (pH 7.1) in the absence
or presence of 4 mM MgCl₂ by CAA, DEPC, or OsO₄ as
described (20). DNA samples were then digested with the
restriction enzyme HindIII, end-labeled, and treated with
piperidine. CAA-modified samples were treated with either
formic acid or hydrazine in high salt prior to piperidine.

Sequencing of Double-Stranded DNA. Three to five micro-
grams of supercoiled DNAs containing sequences I and II
was mixed with 0.5 pmol of either "reverse" or "−40"
primers (for synthesis of purine-rich or pyrimidine-rich
strands, respectively), denatured in 200 mM NaOH/0.2 mM
EDTA at 65°C for 10 min, neutralized in 0.75 M sodium
acetate (pH 5.2), and ethanol precipitated. Annealing was
performed in Sequenase buffer (40 mM Tris·HCl, pH 7.5/50

Abbreviations: DEPC, diethyl pyrocarbonate; CAA, chloroacetal-
dehyde.
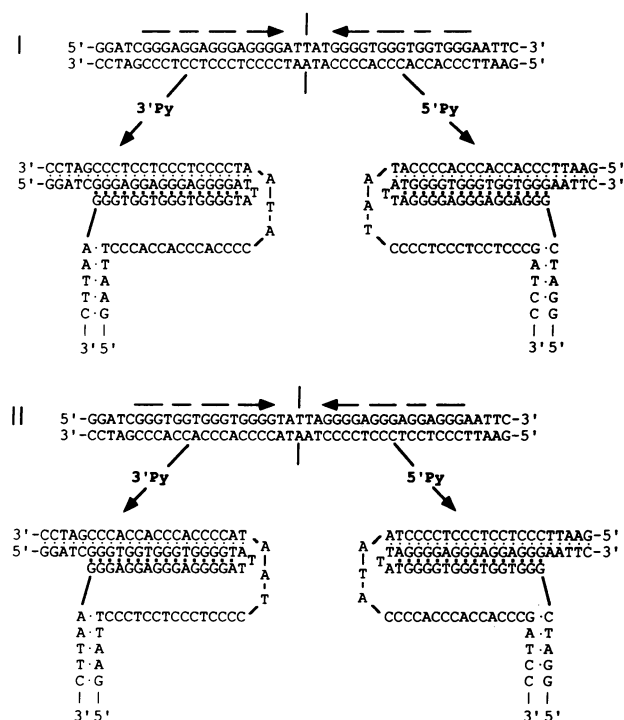[†]To whom reprint requests should be addressed.

FIG. 1.   Intramolecular triplexes consisting of G·G·C and T·A·T triads. In both sequences I and II, G·C base pairs are arranged as mirror images (shown by arrows; the pseudosymmetry axis is shown by a vertical line), whereas A·T base pairs are arranged as inverted repeats. •, Watson–Crick hydrogen bonds; ■, Hoogsteen hydrogen bonds.

mM NaCl/20 mM MgCl₂) for 10 min at 42°C. DNA sequencing was then carried out according to the Sequenase version 2.0 sequencing protocol (United States Biochemical).

Sequencing of Single-Stranded DNA. One-half microgram of single-stranded DNAs containing either the purine-rich or the pyrimidine-rich strand of sequence II was mixed with 15 pmol of the reverse primer. Annealing was performed in Sequenase buffer for 15 min at 37°C. DNA sequencing was then carried out according to the Sequenase version 2.0 sequencing protocol with minor modifications.

RESULTS

DNA Triplexes Formed by Sequences That Are Neither Homopurine–Homopyrimidine nor Mirror Repeats. As described above, to form an intramolecular DNA triplex consisting of G·G·C and T·A·T base triads, a sequence should conform to unusual sequence requirements (Fig. 1). In the purine-rich strand, guanine residues are arranged as a mirror repeat. Guanine clusters are intervened by adenine residues in one half of the mirror repeat, and thymine residues are present in its other half. The apparent asymmetric character of such sequences leads to the existence of two subclasses with adenine residues either in the 5' part or in the 3' part of the mirror repeats.

To test whether such triplexes can be formed, we have cloned two sequences corresponding to the above subclasses (I and II in Fig. 1) into the plasmid pUC19. Supercoiled plasmid DNAs containing these inserts were treated with several chemical agents that are used as tools for detecting unusual DNA conformations, including intramolecular triplexes. These agents include CAA, OsO₄, and DEPC (reviewed in ref. 2). It was shown previously that these chemicals react preferentially with single-stranded DNA segments and junctions between double- and triple-helical DNA within intramolecular triplexes (6–11).

Chemical modifications were carried out at neutral pH in the absence (control) or presence of magnesium ions, which stabilize H*-DNA (11–14). Fig. 2 shows the results of various chemical modification experiments. One can see that in the presence of magnesium ions both sequences show prominent chemical hyperreactivity, consistent with a structural transition. The transition is supercoil-dependent, as we were unable to detect any chemical modification in linear DNA (data not shown). The modification pattern of both sequences differs with respect to the DNA strand. In the purine-rich strand, only the central portion is modified. In the pyrimidine-rich strand, prominent chemical hyperreactivity in the 5' half (for sequence I) or the 3' half (for sequence II) was observed. The modification of half of the pyrimidine strand and the center of the purine strand is consistent with the model of intramolecular DNA triplexes as presented in Fig. 1. We believe, therefore, that our sequences form H*-type triplexes.

For homopurine–homopyrimidine mirror repeats, two isoforms of intramolecular triple helices identical in base triads but different in strand polarities and topological properties could exist in principle (5, 21). Usually, however, the isoform with the 3' part of the purine strand for H-DNA or the isoform with the 3' part of the pyrimidine strand (3'Py) for H*-DNA involved in triplex formation is favored (6–12). Our case is unusual, because the two isoforms of an H* triplex would also differ in base triad composition due to the asymmetric character of the sequences (see Fig. 1). The common triad for both isoforms is G·G·C. For sequence I, the 3'Py isoform contains T·A·T triads, while the isoform with the 5' part of the pyrimidine strand (5'Py) contains A·T·A triads. Conversely, for sequence II, the 3'Py isoform contains A·T·A triads, while the 5'Py isoform contains T·A·T triads. The data in Fig. 2 show that sequences I and II actually form different isoforms of H* triplex: 3'Py and 5'Py, respectively. In both cases, an intramolecular triplex is formed by G·G·C and T·A·T base triads.

DNA Triplexes Block DNA Polymerization. As discussed above, H*-type DNA triplexes are stabilized under conditions close to optimal for DNA polymerase activity *in vitro* (i.e., neutral pH and in the presence of magnesium ions). Therefore, they may appear in the process of DNA polymerization on double-helical DNA templates *in vitro*. To study the possible influence of these triplexes on DNA polymerization, supercoiled plasmid DNAs containing either sequence I or sequence II were alkali denatured (i.e., all DNA secondary structures were destroyed), and then primers were annealed in the presence of magnesium ions. Thus, renaturation conditions were optimal for the formation of H*-DNA. DNA sequencing reactions were then carried out by using the Sanger method with Sequenase version 2.0. The results presented in Fig. 3 show that for both DNAs there are prominent stop points in DNA synthesis, which are reflected by strong bands at the same positions in all four sequencing ladders. However, the exact locations of the termination points differ between sequences I and II. When synthesizing the pyrimidine-rich strand, DNA polymerase stops at the borders between vector DNA and the insert for both sequences. In contrast, when the purine-rich strand is synthesized, the major stop points are located at the border of vector DNA and insert I or at the middle of insert II.

The appearance of stop signals is dependent on renaturation conditions. If we annealed primers in the absence of magnesium, and magnesium was added together with DNA polymerase, the intensity of stop signals decreased to almost background levels (Fig. 4). This shows that the termination of DNA polymerization is not sequence dependent, but depends on the overall structure of the cloned DNA sequences.

We suggest, therefore, that the appearance of termination sites may be due to the reformation of intramolecular DNA
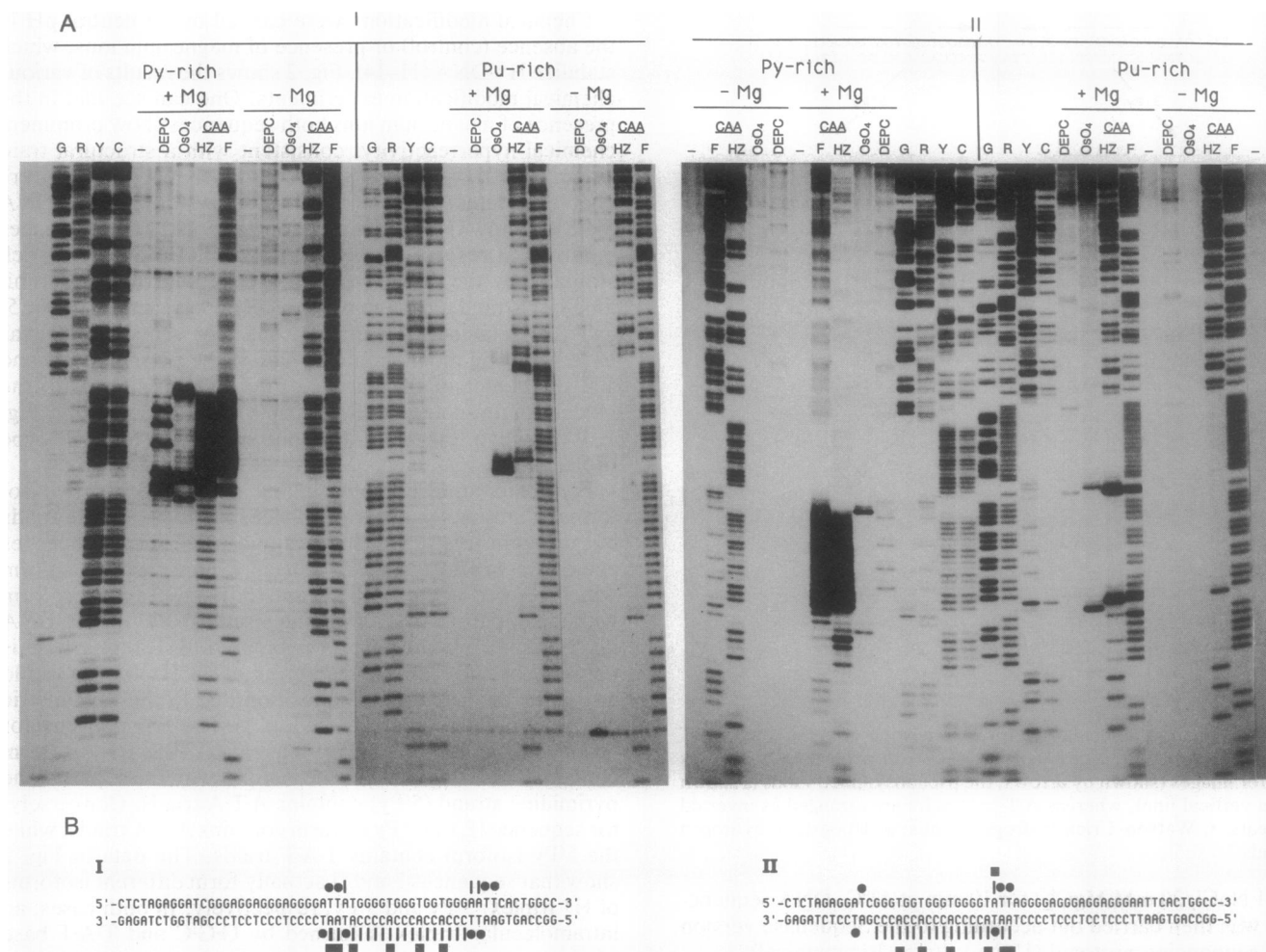
FIG. 2. Chemical modification of sequences I and II in supercoiled DNA. (*A*) Samples were treated in the presence (+ Mg) or absence (− Mg) of magnesium ions with no chemicals (−), DEPC, OsO₄, and CAA (F, formic acid treatment prior to piperidine cleavage; HZ, hydrazine treatment prior to piperidine cleavage). G, R, Y, and C are standard Maxam–Gilbert sequencing ladders for guanine, purine, pyrimidine, and cytosine residues, respectively. (*B*) DNA bases in sequences I and II hyperreactive toward single-stranded DNA-specific chemicals: ■, DEPC; ●, OsO₄; |, CAA. Py-rich, pyrimidine-rich; Pu-rich, purine-rich.

triplexes during template denaturing–renaturing. In support of this idea, the differences in the location of termination sites between sequences I and II could be explained by differences in strand polarity between 3′Py and 5′Py isoforms of triplexes (Fig. 5). Indeed, if the 3′Py isoform of a triplex is formed (sequence I), DNA polymerization should stop at the triplex junction located at the 3′ ends of both strands of the insert. In the case of the 5′Py isoform (sequence II), DNA polymerization should stop at the junction located in the middle of the pyrimidine strand or at the 3′ end of the purine strand.

When we alkali denature covalently closed circular DNA, the following renaturation is difficult to complete due to the topological limitations in DNA chain rotation. Therefore, this template could be considered as a coil of multiple intertwined single-stranded DNA segments, rather than regular double-helical DNA. Thus, one can suggest that the observed termination patterns are due to the structural peculiarities of cloned DNA sequences in a single-stranded state. To check this, we obtained circular single-stranded DNAs containing either the purine-rich or pyrimidine-rich strand of insert II. We compared DNA sequencing of these DNAs in a single- and double-stranded state. The results are presented in Fig. 6. One can see that termination signals appear at similar positions in both single- and double-stranded DNA when the purine-rich strand serves as a template. Thus, the structure of this strand may cause DNA polymerase to terminate. In contrast, when the pyrimidine-rich strand serves as a tem-

plate, we do not see termination signals in single-stranded DNA but detect prominent stops in double-stranded DNA. Therefore, the termination in the pyrimidine-rich template is not due to its structure but requires the presence of the complementary DNA chain. We believe, therefore, that at least in the latter case, intramolecular triplexes formed prior to DNA replication are responsible for premature termination.

DISCUSSION

Our data show that H*-type DNA triplexes can be formed by sequences other than homopurine–homopyrimidine mirror repeats in supercoiled DNA in the presence of bivalent cations. Because of the asymmetrical character of such sequences, different triplex isoforms are distinguished by not only DNA chain polarity but also base triad composition. In our case they may contain either T·A·T or A·T·A triads. Actually, the only triplexes we detected were formed by G·G·C and T·A·T base triads. This shows that the incorporation of the A·T·A triad into a triplex is less likely than the incorporation of the T·A·T triad. As a result, usually the unfavorable 5′Py isoform of H*-DNA becomes favorable for certain DNA sequence (sequence II in Fig. 1). Previously, the formation of this isoform was observed only transiently and at a low level of DNA supercoiling (21). We believe that this is the first case when the preferential formation of the
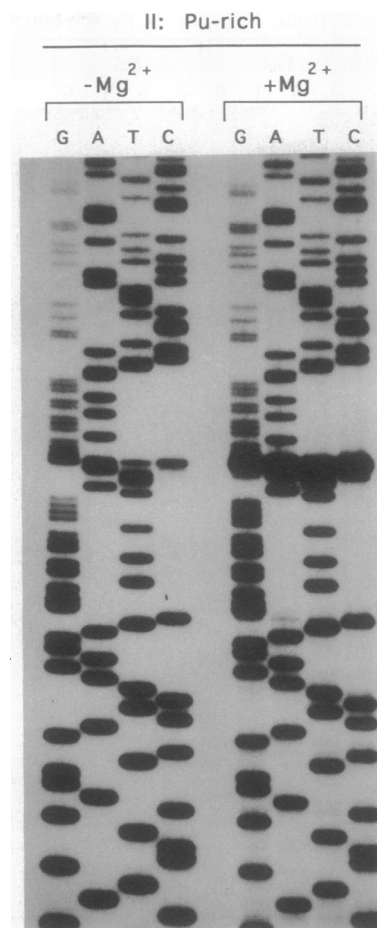
Biochemistry: Dayn *et al.*

Proc. Natl. Acad. Sci. USA 89 (1992) 11409
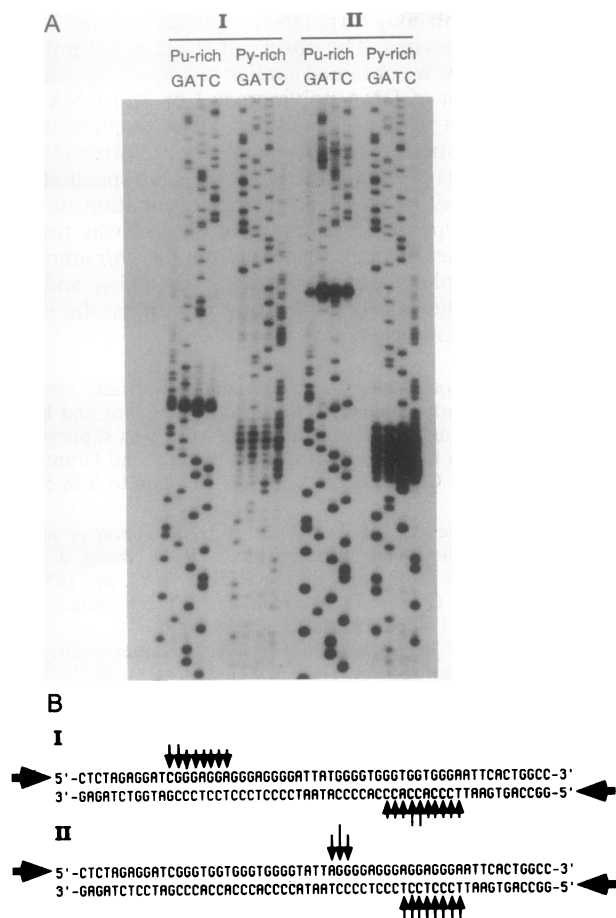
FIG. 3. The influence of triplex formation on DNA polymerization. (*A*) DNA synthesis of the purine (Pu)-rich and the pyrimidine (Py)-rich strands of cloned sequences I and II. G, A, T, and C are ddGTP, ddATP, ddTTP, and ddCTP termination reactions, respectively. (*B*) DNA polymerase stop sites. Horizontal arrows indicate the direction of primer extension in the newly synthesized strands. Vertical arrows show termination sites represented by strong bands in all four sequencing ladders.

FIG. 4. The influence of magnesium ions on DNA polymerase blockage. DNA synthesis of the purine (Pu)-rich strand of cloned sequence II. Denatured plasmid DNA was annealed with the primer either in the absence (*Left*) or in the presence (*Right*) of 20 mM MgCl₂. In the first case, MgCl₂ was added up to 14 mM at the labeling step of DNA sequencing. G, A, T, and C are as in Fig. 3.

energetically unfavorable isoform was enforced by the design of the DNA sequence.

It was found before that H*-DNA may consist of G·G·C and A·A·T triads (11–14). Here we show that the T·A·T base triad may be incorporated in such triplexes as well. The fact that at least three types of base triplets can be involved in H* triplex formation significantly extends the sequence repertoire for triplex formation, thus increasing potential biological applications.

We found that Sanger sequencing of plasmids with the triplex-forming inserts shows distinct termination signals at specific DNA bases. We suggest that this is due to the formation of DNA triplexes prior to DNA polymerization. Three kinds of evidence support this suggestion. First, the termination sites coincide surprisingly well with the triplex edges as defined by chemical probing. Different sequences we studied form isoforms of H*-DNA with different chain polarity. As a result, the processing DNA polymerase would face a triple-helical structure either at the ends or at the middle of the insert. As one can see from Figs. 3 and 5, this is indeed the case, and the stop signals fit precisely with the triplex borders for different isoforms. Second, the appearance of stop signals depends on denaturing–renaturing conditions: omitting magnesium ions in the renaturing mixture causes them to disappear. We also found that triplexes are reformed when denatured plasmids renature in the presence

of magnesium ions (data not shown). This indicates a link between triplex formation and termination of polymerization. Third, when comparing DNA sequencing of single- and double-stranded DNA templates, we observed DNA polymerase stops only in double-stranded DNA in the case of the
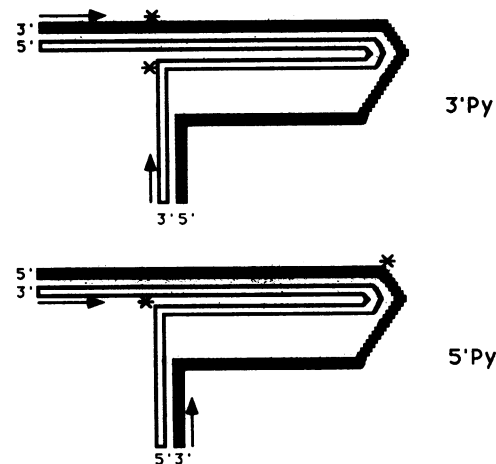
FIG. 5. Model of the termination of DNA polymerization at triplex junctions. ▬, pyrimidine-rich strand; ▭, purine-rich strand. Growing DNA chains are shown by arrows. Asterisks show potential termination sites.
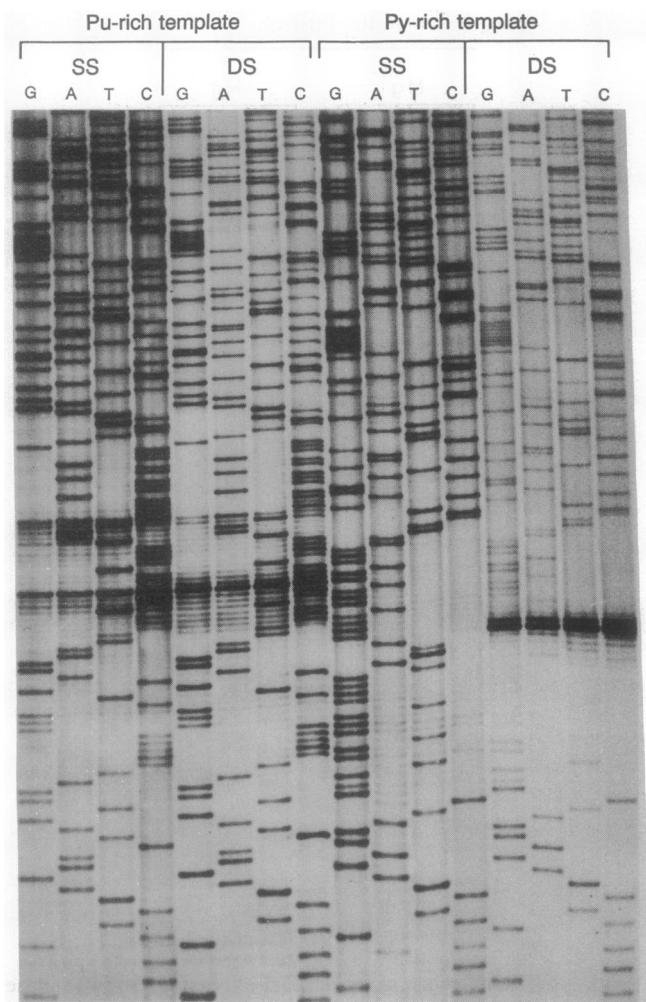
FIG. 6. DNA polymerization on single-stranded (SS) and double-stranded (DS) DNA templates. Two double-stranded DNA represent derivatives of pBluescript SK(−) with sequence II inserted in two orientations. Single-stranded DNAs rescued from the corresponding plasmids contain either the purine (Pu)-rich or the pyrimidine (Py)-rich strand of this insert. G, A, T, and C are as in Fig. 3.

purine-rich strand synthesis. This shows that the existence of both DNA strands in the template, which is a fundamental requirement for triplex formation, accounts for the observed termination.

It is important to note that synthesis on the purine-rich template was terminated even if it was single stranded. Previously, similar observations were made for the (dG-dA)$_n$ inserts in single-stranded DNA (18). In that case, the authors observed termination at the middle of the insert. They explained their data through the formation of a triplex in which a portion of a daughter DNA chain was involved. In our case, we observe most prominent signals at the 3' half of the insert, though we still see minor termination sites in its middle. This difference could be due to the fact that our sequences are much more guanine-rich. In solution, guanine-rich single-stranded DNA pieces can form complex secondary structures stabilized by G·G contacts including hairpins and quadruplexes (22–26). Thus, we assume that the observed termination on a single-stranded template is due to hairpin formation, whereas triplexes combined from these hairpins and segments of the pyrimidine-rich strands cause the same result in a double-stranded template.

In summary, we found that certain G+C-rich DNA sequences that differ from perfect homopurine–homopyrimi-

dine mirror repeats may form DNA triplexes and block DNA polymerization *in vitro*. The observed effect is not polymerase specific, since we obtained similar results using either the Klenow fragment of DNA polymerase I or *Taq* DNA polymerase (data not shown). G+C-rich DNA sequences with perfect and near-perfect mirror symmetry are often found in regulatory regions of eukaryotic DNA (27). We speculate that these repeats may be involved in the termination of DNA replication *in vivo*. In support of this idea, it was recently found that a cluster of simple repeats from the *dhfr* amplicon, including the triplex-forming stretches (dA-dG)$_{21}$ and (dA-dG)$_{27}$, reduces the rate of replication driven by the simian virus 40 replication origin (28).

1. Frank-Kamenetskii, M. D. (1990) in *DNA Topology and Its Biological Effects*, eds. Cozzarelli, N. R. & Wang, J. (Cold Spring Harbor Lab., Cold Spring Harbor, NY), pp. 185–215.
2. Wells, R. D., Collier, D. A., Hanvey, J. C., Shimizu, M. & Wohlrab, F. (1988) *FASEB J.* **2,** 2939–2949.
3. Lyamichev, V. I., Mirkin, S. M. & Frank-Kamenetskii, M. D. (1985) *J. Biomol. Struct. Dyn.* **3,** 327–338.
4. Lyamichev, V. I., Mirkin, S. M. & Frank-Kamenetskii, M. D. (1986) *J. Biomol. Struct. Dyn.* **3,** 667–669.
5. Mirkin, S. M., Lyamichev, V. I., Drushlyak, K. N., Dobrynin, V. M., Filippov, S. A. & Frank-Kamenetskii, M. D. (1987) *Nature (London)* **330,** 495–497.
6. Hanvey, J. C., Klysik, J. & Wells, R. D. (1988) *J. Biol. Chem.* **263,** 7386–7396.
7. Johnston, B. H. (1988) *Science* **241,** 1800–1804.
8. Htun, H. & Dahlberg, J. E. (1988) *Science* **241,** 1791–1796.
9. Voloshin, O. N., Mirkin, S. M., Lyamichev, V. I., Belotserkovskii, B. P. & Frank-Kamenetskii, M. D. (1988) *Nature (London)* **333,** 475–476.
10. Vojtiskova, M., Mirkin, S., Lyamichev, V., Voloshin, O., Frank-Kamenetskii, M. & Palecek, E. (1988) *FEBS Lett.* **234,** 295–299.
11. Kohwi, Y. & Kohwi-Shigematsu, T. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 3781–3785.
12. Kohwi, Y. (1989) *Nucleic Acids Res.* **17,** 4493–4502.
13. Bernües, J., Beltran, R., Casasnovas, J. M. & Azorin, F. (1989) *EMBO J.* **8,** 2087–2094.
14. Bernues, J., Beltran, R. & Azorin, F. (1990) *Nucleic Acids Res.* **18,** 4067–4073.
15. Belotserkovskii, B. P., Veselkov, A. G., Filippov, S. A., Dobrynin, V. N., Mirkin, S. M. & Frank-Kamenetskii, M. D. (1990) *Nucleic Acids Res.* **18,** 6621–6624.
16. Beal, P. A. & Dervan, P. B. (1991) *Science* **251,** 1360–1363.
17. Lapidot, A., Baran, N. & Manor, H. (1989) *Nucleic Acids Res.* **17,** 883–900.
18. Baran, N., Lapidot, A. & Manor, H. (1991) *Proc. Natl. Acad. Sci. USA* **88,** 507–511.
19. Katayama, C. (1990) *Strategies* **4,** 56–58.
20. Pestov, D. G., Dayn, A., Siyanova, E. Y., George, D. L. & Mirkin, S. M. (1991) *Nucleic Acids Res.* **19,** 6527–6532.
21. Htun, H. & Dahlberg, J. E. (1989) *Science* **243,** 1571–1576.
22. Henderson, E., Hardin, C. C., Walk, S. K., Tinoco, I. & Blackburn, E. H. (1987) *Cell* **51,** 899–908.
23. Williamson, J. R., Raghuraman, M. K. & Cech, T. R. (1990) *Cell* **59,** 871–880.
24. Sundquist, W. I. & Klug, A. (1990) *Nature (London)* **342,** 825–829.
25. Panyutin, I. G., Kovalsky, O. I., Budowsky, E. I., Dickerson, R. E., Rikhirev, M. E. & Lipanov, A. A. (1990) *Proc. Natl. Acad. Sci. USA* **87,** 867–870.
26. Smith, F. W. & Feigon, J. (1992) *Nature (London)* **356,** 164–168.
27. Trifonov, E. N. & Brendel, V. (1987) *Gnomic: A Dictionary of Genetic Codes* (VCH, New York).
28. Brinton, B. T., Caddle, M. S. & Heintz, N. H. (1991) *J. Biol. Chem.* **266,** 5153–5161.